2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

FCM-type Fuzzy Co-clustering by K-L Information Regularization

Katsuhiro Honda, Shunnya Oshio and Akira Notsu

***Abstract*— Fuzzy c-Means (FCM) clustering by entropy-based regularization concept is a fuzzy variant of Gaussian mixtures density estimation. FCM was also extended to a full-parameter model by introducing Mahalanobis distance and the K-L information-based fuzzification scheme, in which the degree of fuzziness of partition is evaluated comparing with Gaussian mixtures. In this paper, a new fuzzy co-clustering model is proposed, which is a fuzzy variant of multinomial mixture density estimation. Multinomial mixtures is a probabilistic model for co-clustering of cooccurrence matrices and the proposed method extends multinomial mixtures so that the degree of fuzziness can be tuned in a similar manner to K-L information-based FCM. Several experimental results demonstrate the effects of tuning the degree of fuzziness comparing with its corresponding probabilistic model.**

I. Introduction

FUZZY c-Means (FCM) [1] is the fuzzy extension of the k-Means algorithm [2], where the k-Means objective function is non-linearized by introducing the weighting exponent on fuzzy memberships, because the linear objective function of k-Means can give only crisp membership assignments. A larger weighting exponent gives a fuzzier partition while the model is reduced to crisp k-Means when it equals to one. Although the fuzzy partition can often contribute to effectively revealing intrinsic cluster structures of multivariate observations, the degree of fuzziness is evaluated only from the empirical view points because the standard FCM has no comparative model. (In the following, the standard FCM is represented as sFCM for convenience.)

Non-linear objective function of FCM can also be proposed based on regularization concepts. The entropy-based regularization approach (eFCM) [3], [4] modified the k-Means objective function into non-linear one by adding an entropy-like term with a regularization weight. The weight parameter plays a similar role to the weighting exponent in sFCM. Although the entropy-based objective function was constructed based on fuzzification of k-Means objective function, it can be identified with the negative log-likelihood function of a constrained Gaussian mixture models (GMMs) with spherical covariances. FCM by Kullback-Leibler (K-L) information-based regularization (KLFCM) [5], [6] is a fuzzy variant of the full-parameter GMMs, where model parameters include fuzzy memberships, cluster centers, cluster volumes and full covariance matrices. In the entropy-based and K-L information-based methods, the degree of fuzziness can be evaluated comparing with its probabilistic counterpart of GMMs. Additionally, the regularization concept is also useful for implementing deterministic annealing process of soft clustering [7].

In this paper, a new fuzzy co-clustering algorithm is proposed by modifying a probabilistic mixture models concept. The goal of co-clustering is to extract object-item pairwise structures from cooccurrence matrices. Fuzzy clustering for categorical multivariate data (FCCM) [8] is an FCM-type fuzzy co-clustering algorithm, in which the clustering criterion was defined by the degree of co-cluster aggregations of objects and items. In FCCM, co-cluster structures are represented by using two types of fuzzy memberships of *object memberships* and *item memberships*, and they were fuzzified by the entropy-based regularization concept. Although the FCCM objective function is based on a similar concept to FCM by entropy-based regularization, there is no comparative probabilistic model. Then, in this paper, the proposed algorithm is constructed as a fuzzy counterpart of a probabilistic model.

Multinomial mixture models (MMMs) [9] is a statistical mixture models for discrete distributions, in which component distributions are given by multinomial distributions. The multinomial distribution gives the probability of any particular combination of numbers of successes for the various categories. MMMs has been applied to such a problem as document analysis, where many documents are partitioned into several themes (clusters) and each theme is characterized by the keyword typicalities (category histogram). The document assignment and keyword typicalities are iteratively estimated based on the EM algorithm [10]. The proposed objective function is defined by the MMMs-based log-likelihood function, which includes a K-L information-like regularization term.

The remaining parts of this paper are organized as follows: Section II gives a brief review on the FCM algorithms based on regularization concepts. A new fuzzy co-clustering algorithm is proposed following a brief introduction of MMMs in Section III. The characteristics of the proposed algorithm are demonstrated in Section IV and the summary conclusions are given in Section V.

II. FCM Algorithms Based on Regularization Concepts

A. Fuzzification Mechanisms in FCM

The goal of sFCM [1] is to partition n samples having m-dimensional observation $\boldsymbol{x}_i$, $i = 1,\dots,n$ into C fuzzy clusters, whose prototypes are their prototypical centroids $\boldsymbol{b}_c$, $c = 1,\dots,C$. The clustering criteria are given by the

Katsuhiro Honda, Shunnya Oshio and Akira Notsu are with the Graduate School of Engineering, Osaka Prefecture University, Osaka, Japan (email: {honda, notsu}@cs.osakafu-u.ac.jp).

This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology, Japan, through a Grant-in-Aid for Scientific Research (#23500283, #26330281).

within-cluster errors as:

$$L_{sfcm} = \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci}^{\theta}||\boldsymbol{x}_i - \boldsymbol{b}_c||^2, \quad (1)$$

where u_{ci} is the membership of sample i to cluster c, and is normalized as $\sum_{c=1}^{C} u_{ci} = 1$. L_{sfcm} is a fuzzified version of k-Means objective function [2], where the linear k-Means function is non-linearized by introducing the exponential weight θ ($\theta > 1$).

eFCM [3], [4] adopted another fuzzification mechanism, in which k-Means function is non-linearized by adding non-linear regularization terms.

$$L_{efcm} = \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci}||\boldsymbol{x}_i - \boldsymbol{b}_c||^2 + \lambda \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci}\log u_{ci}. \quad (2)$$

The additional negative entropy term, which is a downward-convex function, forces u_{ci} to have fuzzy memberships while the crisp k-Means function is minimized in $u_{ci} = \{0, 1\}$. λ tunes the degree of fuzziness of partition, and the larger λ is, the fuzzier the partition is.

Although the fuzzy clustering model is constructed as a fuzzy version of k-Means, the objective function has a close connection with the negative log-likelihood function of a constrained GMMs with spherical covariances [11]. Assume that component densities of GMMs, which is composed of C components, are given as:

$$g_c(\boldsymbol{x}_i|\boldsymbol{b}_c) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{||\boldsymbol{x}_i - \boldsymbol{b}_c||^2}{2\sigma^2}\right), \quad (3)$$

and they are combined into a mixture density as:

$$P(\boldsymbol{x}_i) = \frac{1}{C}\sum_{c=1}^{C} g_c(\boldsymbol{x}_i|\boldsymbol{b}_c). \quad (4)$$

This is a spherical Gaussian model, where the component covariance is σI with a fixed σ and is common for all components. The a priori probabilities of components are also fixed and common as $1/C$. Then, the lower-bound of the negative log-likelihood function to be minimized is reduced to Eq.(2), where u_{ci} is the posteriori probability of component c given by sample i. λ corresponds to $2\sigma^2$, i.e., the component covariance is pre-fixed and is not updated.

KLFCM [5], [6] is a fuzzy variant of the full-parameter GMMs, in which full-elements of covariance matrices and the cluster volumes (a priori probability of components) are also updated. In this paper, for simplicity, a spherical Gaussian model is considered. KLFCM introduced an additional cluster volume parameter α_c, which can be identified with the mixing coefficient of each component density in GMMs, and used the following objective function supported by K-L information-based fuzzification mechanism:

$$L_{klfcm} = \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci}||\boldsymbol{x}_i - \boldsymbol{b}_c||^2 - \lambda \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci}\log\frac{\alpha_c}{u_{ci}}. \quad (5)$$

The optimal α_c with fixed u_{ci} is derived as

$$\alpha_c = \frac{1}{n}\sum_{i=1}^{n} u_{ci}, \quad (6)$$

and the K-L information term, whose maximum is 0 for $u_{ci} = \alpha_c$, is larger as u_{ci} become more similar to α_c and more homogeneous in each cluster. In this sense, memberships are fuzzified reflecting the cluster volumes in each cluster.

KLFCM is also reduced to GMMs in the case where λ is equivalent to a double within-cluster variance, but has no identical statistical models in other cases.

B. Fuzzy Co-clustering Models

In fuzzy co-clustering contexts, a different type of clustering criteria was adopted, in which mutually familiar objects and items are merged into co-clusters considering the aggregation degree of each co-cluster. Assume that we have $n \times m$ cooccurrence information $R = \{r_{ij}\}$ among n objects and m items, and the goal is to simultaneously estimate fuzzy memberships of objects u_{ci} and items w_{cj}. The sum of aggregation degrees to be maximized is defined as:

$$L = \sum_{c=1}^{C}\sum_{i=1}^{n}\sum_{j=1}^{m} u_{ci}w_{cj}r_{ij}. \quad (7)$$

In order to avoid trivial solutions, w_{cj} are forced to be exclusive in each cluster such that $\sum_{j=1}^{m} w_{cj} = 1$ while u_{ci} are estimated under the same condition with FCM such that $\sum_{c=1}^{C} u_{ci} = 1$. Then, w_{cj} represent the relative typicalities of items in each cluster.

Considering the maximization nature, the two types of memberships are fuzzified based on regularization concepts instead of the standard exponential-type weighting, which is designed for minimization principles. FCCM [8] used the entropy-based fuzzification [3], [4]:

$$\begin{aligned} L_{fccm} \;=\; & \sum_{c=1}^{C}\sum_{i=1}^{n}\sum_{j=1}^{m} u_{ci}w_{cj}r_{ij} \\ & -\lambda_u \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci}\log u_{ci} \\ & -\lambda_w \sum_{c=1}^{C}\sum_{j=1}^{m} w_{cj}\log w_{cj}, \end{aligned} \quad (8)$$

where λ_u and λ_w are the fuzzification weights for object and item memberships, respectively. The entropy terms, which are convex functions, work for fuzzifying memberships. Kummamuru *et al.* [12] extended FCCM by introducing the quadric term-based fuzzification mechanism [13].

Although u_{ci} and w_{cj} have a similar partition concept with probabilistic mixture models, there is no counterparts of these fuzzy co-clustering models, i.e., we have no measure for evaluating the degree of fuzziness comparing with probabilistic models.

The remaining part of this paper discusses the connection with fuzzy co-clustering and probabilistic mixture models, and proposes a novel fuzzy co-clustering algorithm.

III. Fuzzy Co-clustering with K-L Information Regularization Based on Multinomial Mixture Concepts

A. Co-clustering by Multinomial Mixture Models

Assume that $n \times m$ cooccurrence matrix $R = \{r_{ij}\}$ is composed of frequency r_{ij} of item j in object i, where item j is observed r_{ij} times with object i. Multinomial distribution is a multi-category extension of binomial distribution. When all objects were drawn from a single distribution, the probability of observation j drawn from all m items is defined as w_j. The joint distribution of cooccurrence feature vector $\boldsymbol{r}_i = (r_{i1}, \ldots, r_{im})^\top$ on object i is given as:

$$p(\boldsymbol{r}_i) = \frac{T_i!}{r_{i1}! \ldots r_{im}!} \prod_{j=1}^{m} (w_j)^{r_{ij}}, \tag{9}$$

where T_i is the total observation of $T_i = \sum_{j=1}^{m} r_{ij}$. Then, the log-likelihood function on all n objects is

$$\begin{aligned} L_{mm} &= \sum_{i=1}^{n} \log(p(\boldsymbol{r}_i)) \\ &= \sum_{i=1}^{n} \log\left(\frac{T_i!}{r_{i1}! \ldots r_{im}!} \prod_{j=1}^{m} (w_j)^{r_{ij}} \right). \end{aligned} \tag{10}$$

The maximum likelihood solution for w_j is derived as $w_j = t_j / T$, where t_j is the sum of frequency of item j such that $t_j = \sum_{i=1}^{n} r_{ij}$ and T is the total observation such that $T = \sum_{i=1}^{n} T_i$.

MMMs [9] is a statistical model for co-clustering, in which probabilistic mixture models is constructed with multinomial component densities. When objects are assume to be drawn from C different distributions, in which the probability of item observation j from component c is defined as w_{cj}, the mixture distribution is given as:

$$\begin{aligned} P(\boldsymbol{r}_i) &= \sum_{c=1}^{C} \pi_c p_c(\boldsymbol{r}_i) \\ &= \sum_{c=1}^{C} \pi_c \frac{T_i!}{r_{i1}! \ldots r_{im}!} \prod_{j=1}^{m} (w_{cj})^{r_{ij}}. \end{aligned} \tag{11}$$

Then, the log-likelihood function on all n objects is

$$\begin{aligned} L_{mmms} &= \sum_{i=1}^{n} \log(P(\boldsymbol{r}_i)) \\ &= \sum_{i=1}^{n} \log\left(\sum_{c=1}^{C} \pi_c p_c(\boldsymbol{r}_i) \right) \\ &= \sum_{i=1}^{n} \log\left(\sum_{c=1}^{C} u_{ci} \frac{\pi_c p_c(\boldsymbol{r}_i)}{u_{ci}} \right), \end{aligned} \tag{12}$$

where u_{ci} is the posterior probability of component c given object i.

It has been shown that the maximum likelihood solution for Eq.(12) is derived by maximizing the following pseudo-log-likelihood function:

$$\begin{aligned} L'_{mmms} &= \sum_{c=1}^{C} \sum_{i=1}^{n} \sum_{j=1}^{m} u_{ci} r_{ij} \log w_{cj} \\ &\quad + \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} \log \frac{\pi_c}{u_{ci}}. \end{aligned} \tag{13}$$

B. Fuzzy Co-clustering Based on K-L Information Regularization

Considering the similarity between the FCCM objective function of Eq.(8) and the pseudo-log-likelihood function of Eq.(13), a novel fuzzy co-clustering algorithm is proposed in this paper. The novel algorithm is an extension of FCCM, which has the following three modifications:

1) The cluster aggregation is calculated using $\log w_{cj}$ instead of w_{cj}. Comparing with FCCM, the objective function is expected to be quite sensitive for small w_{cj} using $\log$ function.
2) The fuzzification term (entropy term) on w_{cj} is removed. The fuzzy nature of w_{cj} is supported by the non-linearity of $\log$ function, and the degree of fuzziness on item memberships is fixed.
3) An additional parameter π_c of cluster volume is introduced and u_{ci} is fuzzified with K-L information-based regularization approach by optimizing cluster volumes.

The novel objective function is defined as follows:

$$\begin{aligned} L_{klfccm} &= \sum_{c=1}^{C} \sum_{i=1}^{n} \sum_{j=1}^{m} u_{ci} r_{ij} \log w_{cj} \\ &\quad + \lambda_u \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} \log \frac{\pi_c}{u_{ci}}. \end{aligned} \tag{14}$$

In this formulation, u_{ci} and w_{cj} are estimated under the same constraints with FCCM, i.e., $\sum_{c=1}^{C} u_{ci} = 1$ and $\sum_{j=1}^{m} w_{cj} = 1$. λ_u is the fuzzification weight on u_{ci} and the fuzzy partition becomes more fuzzy as λ_u is larger.

Based on the Lagrangean multiplier method, the updating rules for the model parameters are derived as follows:

$$u_{ci} = \frac{\pi_c \prod_{j=1}^{m} (w_{cj})^{r_{ij}/\lambda_u}}{\sum_{\ell=1}^{C} \pi_\ell \prod_{j=1}^{m} (w_{\ell j})^{r_{ij}/\lambda_u}}, \tag{15}$$

$$\pi_c = \frac{1}{n} \sum_{i=1}^{n} u_{ci}, \tag{16}$$

$$w_{cj} = \frac{\sum_{i=1}^{n} r_{ij} u_{ci}}{\sum_{\ell=1}^{m} \sum_{i=1}^{n} r_{i\ell} u_{ci}}. \tag{17}$$

The clustering algorithm is the 3-step iterative process and a sample procedure is given as:

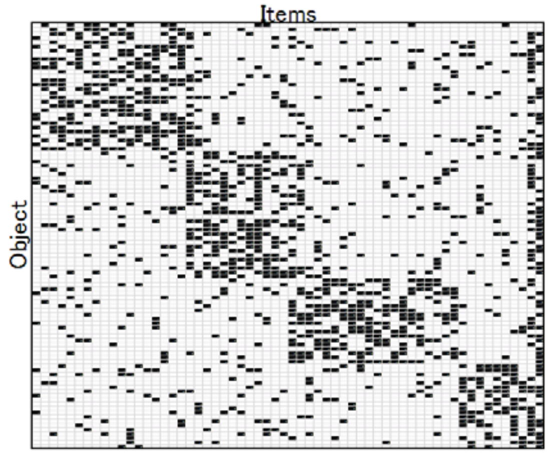


Fig. 1. Artificial data for comparison with MMMs: black and white cells implies $r_{ij} = 1$ and $r_{ij} = 0$, respectively

[Fuzzy Clustering for Categorical Multivariate Data by K-L Information Regularization (KLFCCM)]

1) Let C be the number of clusters. Choose the fuzzification weights λ_u.
2) **[Initialization]** Randomly initialize object memberships u_{ci} and normalize them so that $\sum_{c=1}^{C} u_{ci} = 1$.
3) **[Iterative process]** Iterate the following process until convergence of all u_{ci}.
   a) Update cluster volumes π_c using Eq.(16).
   b) Update item memberships w_{cj} using Eq.(17).
   c) Update object memberships u_{ci} using Eq.(15).

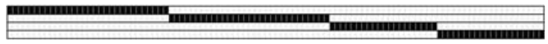
This algorithm is reduced to the conventional MMMs if the fuzzification weight is $\lambda_u = 1$. There is, however, no probabilistic counterpart when $\lambda_u \neq 1$, i.e., the proposed algorithm has an advantage of tuning the fuzzy degree of object partition.

IV. Numerical Experiments

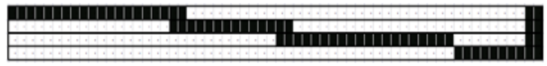
In this section, two experimental results are shown for demonstrating the characteristic features of the proposed method.

A. Comparison with MMMs

First, the characteristics of the proposed method are compared with MMMs. The proposed KLFCCM model can tune the degree of fuzziness of object memberships while $\lambda_u = 1$ reduces it to the conventional MMMs. In this experiment, the effect of the additional fuzzification weight is investigated. A comparative experiment was performed with an artificial data set composed of 100 objects and 60 items. The 100×60 cooccurrence matrix is shown in Fig. 1 by gray-scale (black and white are for $r_{ij} = 1$ and $r_{ij} = 0$, respectively), which includes roughly 4 co-clusters in diagonal blocks while some items are shared by multiple clusters. The ideal object and item memberships of 4 co-clusters are depicted as Figs. 2-(a) and (b), in which each row shows the 100-dimensional object membership

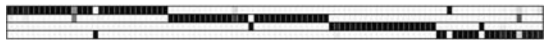


(a) Ideal object membership vectors $\boldsymbol{u}_c$



(b) Ideal item membership vectors $\boldsymbol{w}_c$

(c) Derived object membership vectors $\boldsymbol{u}_c$ ($\lambda_u = 1.0$)



(d) Derived object membership vectors $\boldsymbol{w}_c$ ($\lambda_u = 1.5$)

Fig. 2. Comparison of memberships (KLFCCM v.s. MMMs)

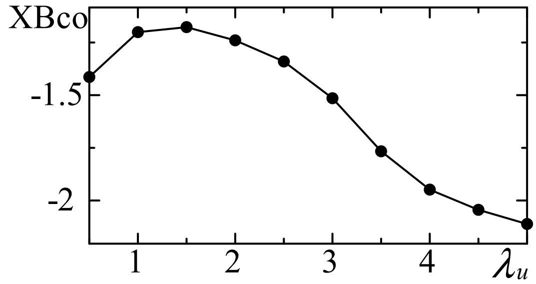


Fig. 3. Comparison of XBco validity measures (KLFCCM v.s. MMMs)

vector $\boldsymbol{u}_c = (u_{c1}, \ldots, u_{c,100})^\top$ or the 60-dimensional item membership vector $\boldsymbol{w}_c = (w_{c1}, \ldots, w_{c,60})^\top$ by gray-scale (black and white are for w_{max} and 0, respectively), and the goal is to extract a similar structure from the noisy data set.

Figures 2-(c) and (d) compare the object memberships derived by MMMs ($\lambda_u = 1$) and the proposed method with $\lambda_u = 1.5$. Because a fuzzier partition is often robust against noise, a slightly fuzzier model of $\lambda_u = 1.5$ could derive a better result.

In order to validate the intuitive recognition of Fig. 2, the quality of co-cluster partitions are evaluated by using a quantitative measure. XBco [14] is a Xie-Beni-type validity measure [15] for fuzzy co-cluster partitions, which tries to select the best compact-separate clusters. A larger XBco implies a better co-cluster partition.

$$
\begin{aligned}
XBco &= \frac{compactness}{separateness} \\
&= \frac{(C-1)\displaystyle\sum_{c=1}^{C}\sum_{i=1}^{n}\sum_{j=1}^{m} u_{ci} w_{cj} (2r_{ij} - 1)}{\displaystyle\sum_{k=1}^{C}\sum_{\ell \neq k}\sum_{i=1}^{n}\sum_{j=1}^{m} u_{ki} w_{\ell j} r_{ij}}. \quad (18)
\end{aligned}
$$

Figure 3 compares the XBco values for several different co-cluster partitions derived by the proposed method with different fuzzy weights λ_u. The figure implies that the MMMs partition given by $\lambda_u = 1$ is inferior to the KLFCCM partition given by $\lambda_u = 1.5$, which is slightly fuzzier. However, much more or much less fuzzier models could derive poor results only. By carefully tuning fuzzy degrees,

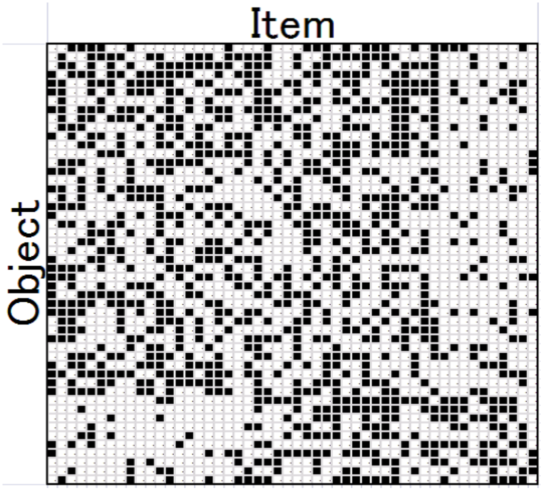


Fig. 4. Artificial data for comparison with FCCM: black and white cells implies $r_{ij} = 1$ and $r_{ij} = 0$, respectively

(a) Ideal object membership vectors: $(u_{c1}, \ldots, u_{c,50})$

(b) Ideal item membership vector: $(w_{c1}, \ldots, w_{c,50})$

Fig. 5. Ideal object and item membership vectors of Fig. 4

the proposed KLFCCM algorithm can reveal intrinsic co-cluster structures.

B. Comparison with FCCM

Second, the proposed KLFCCM is compared with the conventional FCCM, which is fuzzified by the entropy-based regularization method. This experiment was performed with the 50×50 cooccurrence matrix shown in Fig. 4, which includes 2 rough co-clusters. The ideal object and item membership vectors are depicted in Fig. 5 with gray-scale. The cluster volumes (numbers of objects) are quite unbalanced and many items are shared by the big and small clusters, i.e., the cluster boundary is somewhat ambiguous.

FCCM was implemented with $\lambda_w = 10.0$ and various λ_u, whose XBco validity measures are compared in Fig. 6. The figure implies that $\lambda_u = 0.01$ seems to be plausible, and the object membership vectors are shown in Fig. 7-(a). Comparing with the ideal vectors of Fig. 5-(a), the small cluster (2nd row) illegally exploited some objects from the large cluster (1st row). Because FCCM does not consider the optimization of cluster volumes, the cluster boundary tend to located at the middle of two clusters so that the both cluster volumes are to be homogeneous. Note that we could find a similar phenomena also in the FCM case (FCM v.s. KLFCM) [5].

Next, KLFCCM was performed with various λ_u, whose XBco validity measures are compared in Fig. 8. The figure also implies that a slightly fuzzier model of $\lambda_u = 1.5$ than MMMs ($\lambda_u = 1.0$) should be selected. The derived object membership vectors are compared in Fig. 7 where alternative candidates of $\lambda_u = 1.0$ and $\lambda_u = 2.0$ are also

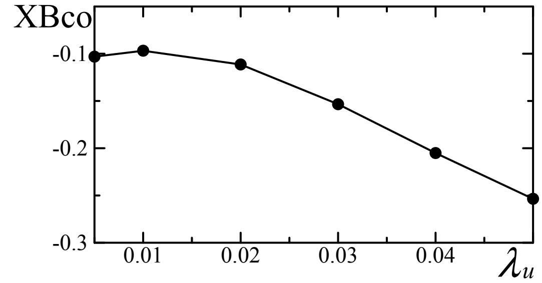


Fig. 6. Comparison of XBco validity measures (FCCM with unbalanced data)

(a) FCCM

(b) KLFCCM ($\lambda = 1.0$)

(c) KLFCCM ($\lambda = 1.5$)

(d) KLFCCM ($\lambda = 2.0$)

Fig. 7. Comparison of object memberships (unbalanced data)

shown for references. The figure implies that, in the results of KLFCCM, the cluster boundary was pushed toward the small cluster and the large cluster tends to have much more objects than FCCM. Because of cluster volume optimization, the majority cluster is emphasized. Especially, the effect of cluster volume parameter was remarkable in fuzzier cases but the selected result of $\lambda_u = 1.5$ still seems to be plausible. (Note that such cluster volume optimization could not be found in FCCM even if the fuzzification weight λ_u was moved.)

V. Conclusions

In this paper, a new algorithm for fuzzy co-clustering was proposed considering the mutual connections between FCCM and MMMs. Introducing the K-L information-based fuzzification mechanism, a better clustering ability of the fuzzy model was demonstrated. The connection with such probabilistic model as MMMs makes it possible to discuss the plausibility of fuzzy degrees comparing with probabilistic counterparts. The advantage of cluster volume optimization was also demonstrated for clarifying the comparison with the conventional FCCM.

A possible future work is to adopt the deterministic annealing approach [7] by exploiting the controllable fuzzification penalty. Introduction of the mechanism for tuning the fuzzy degree of item memberships is also in future work.

References

[1] J. C. Bezdek: *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, 1981.
[2] J. B. MacQueen, "Some methods of classification and analysis of multivariate observations," *Proc. of 5th Berkeley Symposium on Math. Stat. and Prob.*, pp. 281-297, 1967.

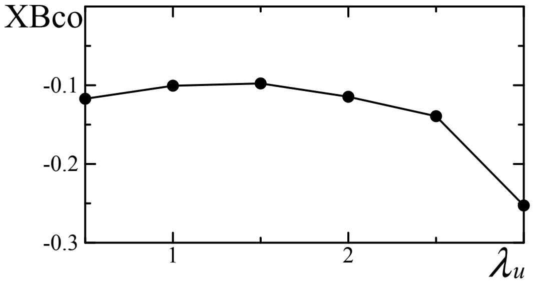


Fig. 8. Comparison of XBco validity measures (KLFCCM with unbalanced data)

[3] S. Miyamoto and M. Mukaidono, "Fuzzy *c*-means as a regularization and maximum entropy approach," *Proc. of the 7th International Fuzzy Systems Association World Congress*, vol. 2, pp. 86-92, 1997.

[4] S. Miyamoto, H. Ichihashi and K. Honda: *Algorithms for Fuzzy Clustering*, Springer, 2008.

[5] H. Ichihashi, K. Miyagishi and K. Honda, "Fuzzy *c*-means clustering with regularization by K-L information," *Proc. of 10th IEEE International Conference on Fuzzy Systems*, vol. 2, pp. 924-927, 2001.

[6] K. Honda and H. Ichihashi, "Regularized linear fuzzy clustering and probabilistic PCA mixture models," *IEEE Trans. Fuzzy Systems*, vol. 13, no. 4, pp. 508-516, 2005.

[7] K. Rose, E. Gurewitz and G. Fox, "A deterministic annealing approach to clustering," *Pattern Recognition Letters*, vol. 11, pp. 589-594, 1990.

[8] C.-H. Oh, K. Honda and H. Ichihashi, "Fuzzy clustering for categorical multivariate data," *Proc. of Joint 9th IFSA World Congress and 20th NAFIPS International Conference*, pp. 2154-2159, 2001.

[9] L. Rigouste, O. Cappé and F. Yvon, "Inference and evaluation of the multinomial mixture model for text clustering," *Information Processing and Management*, vol. 43, no. 5, pp. 1260-1280, 2007.

[10] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. of the Royal Statistical Society*, Series B, vol.39, pp. 1-38, 1977.

[11] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, 1973.

[12] K. Kummamuru, A. Dhawale and R. Krishnapuram, "Fuzzy co-clustering of documents and keywords," *Proc. 2003 IEEE Int'l Conf. Fuzzy Systems*, vol. 2, pp. 772-777, 2003.

[13] S. Miyamoto and K. Umayahara, "Fuzzy clustering by quadratic regularization," *Proc. 1998 IEEE Int. Conf. Fuzzy Systems and IEEE World Congr. Computational Intelligence*, vol. 2, pp. 1394-1399, 1998.

[14] K. Honda, M. Muranishi, A. Notsu, H. Ichihashi, "FCM-type cluster validation in fuzzy co-clustering and collaborative filtering applicability," *International Journal of Computer Science and Network Security*, vol. 13, no. 1, pp. 24-29, 2013.

[15] X. L. Xie, and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.* vol. 13, pp. 841-847, 1991.